SAFEGUARDING QUANTUM KEY DISTRIBUTION THROUGH DETECTION RANDOMIZATION

Thiago Ferreira da Silva[1,2,*], Gustavo C. do Amaral[1], Guilherme B. Xavier[3], Guilherme P. Temporão[1], and Jean Pierre von der Weid[1]

[1] Center for Telecommunication Studies, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.
[2] Optical Metrology Division, National Institute of Metrology, Quality and Technology, Duque de Caxias, RJ, Brazil.
[3] Departamento de Ingeniería Eléctrica/ Centre for Optics and Photonics/ MSI-Nucleus on Advanced Optics, Universidad de Concepción, Concepción, Chile.
[*] thiago@opto.cetuc.puc-rio.br

Quantum key distribution (QKD) allows two remote parties to share a random secret key to be used for private key cryptography. The security of QKD is fully based on the laws of quantum physics and has been unconditionally proven. Based on these assumptions, many experimental – and commercial – systems have been implemented in the last three decades.

Recently, a critical point was recognized: back-doors may be open in some physical devices comprising the QKD system, so *quantum hacking attacks* have been reported to be possible [1,2], specially those aimed at the single-photon detector (SPD). The flaws may be explored by an eavesdropper (Eve) for side-channel attacks, which can jeopardize the security of the protocol. These quantum hacking attacks are interventions caused by Eve from the outside of Bob's station by high-jacking the detection apparatus – whose response can be predicted in some degree or even manipulated. In all cases, the attacks make it possible for an eavesdropper to gain information without being noticed, i.e., achieving a critically high mutual information with Alice and Bob without exceeding the upper threshold of the quantum bit error rate (QBER).

The hacking schemes basically aim at two key points: exploring the imperfect nature of the SPD – efficiency mismatched-based attacks – or externally forcing a deterministic result on the detection equipment – bright-light-based attacks. Different countermeasures to avoid detector-aimed quantum hacking attacks have been presented (see references in [3]). Despite being effective for the proposed specific end, the solutions have no guarantees of being final, in the sense that the vulnerabilities depend on the physical implementation of the devices and the deployment of the systems. The counter-measures give, in the best case, *ad hoc* protection over some class of attacks.

We propose [3] a practical solution that extends over a broad range of known classes of quantum hacking attacks aimed at the detection equipment. The scheme is based on fundamental randomization of input modes to the detection apparatus inside Bob's station, thus not deterministically accessible to the eavesdropper, and comprises two parts: the *spatial modes randomization* and the *detector scrambling*.

We show that the use of beamsplitters and extra detectors at Bob's station in the *spatial modes randomization* strategy renders its apparatus immune to bright-light based attacks, as the blinding- and faked-states attacks. The eavesdropper can no longer manipulate the detectors without leaving a strong signature which is monitored by the counting statistics of the detectors. Correlation between detectors in equivalent spatial modes reveals the attack, without intervening on the inner workings of the devices.

We also emulate the *aftergate attack* [4] against a BB84-like QKD system and show that our method is capable of identifying the attack, even if a most general strategy of mixing faked states and original states is employed by the eavesdropper. We also give the big picture for the degree of protection provided by our proposed counter-measured against bright-light-based attacks when fixed asymmetry of the BSs is considered – assuming that Eve cannot manipulate the splitting ratio of the BSs. The fraction of events stolen by Eve is directly related to the asymmetry caused by the attack and the protection is directly given by the ability of Bob to check the asymmetry out.

The *detector-scrambling* strategy is employed by Bob dynamically. This strategy randomly alters the detector used for measurement under the chosen basis, counteracting the detection efficiency mismatching attacks. We emulated the *time-shift attack* [5], in which Eve controls the time-of-flight of the pulses between Alice and Bob. The countermeasure is equivalent to an active randomization of the spatial modes. When Bob's basis matches Alice's, the logical result is deterministic, but the detector that registered the event is random. We show that the strategy drastically reduces the mutual information between Eve and Bob, dropping significantly towards zero (the ideal value), while no additional hardware is needed. Provided that this choice is truly random, Eve cannot infer at which SPD Bob's detection has occurred, so the final spatial mode cannot be accessed by her. In the time-shift attack context, if Eve waits the basis reconciliation and learns that Alice and Bob agreed, there is no way to infer which version of the basis Bob had chosen and the detector he had used, based on the imposed delay, even if the efficiency curves are fully mismatched. The countermeasure is equivalent to an active randomization of the spatial modes and drastically reduces the mutual information between Alice-Eve and Bob-Eve. When Bob's basis matches Alice's, the logical result is deterministic, but the detector that registered the event is random.

Despite the fact that a full practical solution for all kinds of quantum hacking attacks aimed at all aspects of traditional QKD systems has not yet been found or proved to be possible, the recent overflow of eavesdropping schemes motivated the proposal of many practical solutions. We have shown how some back-doors at the detection end can be closed in a standard BB84 frame with the creation of randomized spatial modes, passively by a combination of beamsplitters and extra SPDs and by actively scrambling the detectors at the measurement station. This represents a practical and readily implementable solution against bright-light- and efficiency-mismatching-based quantum hacking attacks aimed on the single-photon detector so far without tampering on the internals of the detectors.

[1] L. Lydersen *et al.* Nat. Photonics (2010) **4**, 686–689.

[2] I. Gerhardt *et al.*, Nat. Commun. (2011) **2**, 349.

[3] T. Ferreira da Silva *et al.*, IEEE J. Sel. Topics Quantum Electron. (2015) **21**, 1-9.

[4] C. Wiechers *et al.*, New J. Phys. (2011) **13**, 013043.

[5] B. Qi *et al.*, Quantum Inf. Comput. (2007) **7**, 073–082.